Our Ref. No.: 797

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

Express Mail Label No.: EF415382545US

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. BACKGROUND OF THE INVENTION

1.1 TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

1.2 BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

2. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1 – 362 and are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenosine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1 – 362 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1 – 362. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1 – 362 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1 – 362. The sequence

information can be a segment of any one of SEQ ID NO: 1 – 362 that uniquely identifies or represents the sequence information of SEQ ID NO: 1 – 362.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information is provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-362 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-362 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1–362; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1–362; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1–362. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under

stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1–362; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in the Sequence Listing.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-362; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These

techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g., in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form

the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound the binds to a polypeptide of the invention is identified.

The methods of the invention also provides methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products.

Compounds and other substances can effect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 1); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

3. DETAILED DESCRIPTION OF THE INVENTION

3.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only some of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules.

The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonculeotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and N is A, C, G or T (U). It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of

oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NOs:1-362.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NOs: 1-362. The sequence information can be a segment of any one of SEQ ID NOs: 1-362 that uniquely identifies

or represents the sequence information of that sequence of SEQ ID NO: 1-362. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1 \div 4^{25}$) times the increased probability for mismatch at each nucleotide position (3 x 25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol)

and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, *e g.*, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions,

deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g., E. coli*, will be free of

glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted"

proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2):134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligos), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no

more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more that 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 90% sequence identity. Substantially equivalent nucleotide sequences of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, and most preferably at least about 95% identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method (Hein, J. (1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and

the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

3.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1 – 362; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO:1 – 362; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1 - 362. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO: 1 – 362; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 1- 362. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include all of the coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known

methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1 – 362 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1 - 362 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1 - 362 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, more typically at least about 90%, and even more typically at least about 95%, sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1 - 362, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to any one of the polynucleotides of the invention) are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can

differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1 - 362, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NOs: 1 - 362 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology result for the nucleic acids of the present invention, including SEQ ID NOs: 1 - 362, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for Basic Local Alignment Search Tool is used to search for local sequence alignments (Altshul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using Fastxy algorithm.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably

constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by

the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention can be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-362, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and

the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NOs: 1 - 362 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NOs: 1 - 362 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the

protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived

from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intramuscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

3.3 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous

promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7

lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5′ flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli, Bacillus subtilis, Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the

invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this

purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

3.4 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 1-362 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NOs: 1 - 362 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NOs: 1 – 362 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 1-362 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 1-362 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, typically at least about 95%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 1-362.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for

example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments

of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in Molecular Cloning: *A Laboratory Manual*; Ausubel et al., *Current Protocols in*

Molecular Biology. Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 1-362.

The protein of the invention may also be expressed as a product of transgenic animals, *e.g.,* as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, *e.g.,* U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in

the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin

(TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

3.4.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), pFam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobocity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

3.5 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient

expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired

protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting

sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

3.6 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased

protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

3.7 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

3.7.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when

labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology:

Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

3.7.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

3.7.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19;

Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin-γ, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter

6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

3.7.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotential or pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of

mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a

specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering eds.* Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

3.7.5 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation

of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I.

Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

3.7.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue.

De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising

such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

3.7.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes

viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also to be useful in the treatment of allergic reactions and conditions (*e.g.*, anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxocol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent.

Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected

cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

3.7.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

3.7.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

3.7.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostatis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

3.7.11 CANCER DIAGNOSIS AND THERAPY

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases,

blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention (including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl,

Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

3.7.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of

such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of

colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

3.7.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science 282*:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

3.7.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind

polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.*, increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications i.e. phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

3.7.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or

promote an inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflamation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic mylegenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

3.7.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

3.7.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases

or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple

sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive

bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

3.7.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

3.7.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving

inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

3.7.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et at., 1983, Science, 219:56, or by B. Waksman et al., 1963,

Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

3.8 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

3.8.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically,

the amount of polypeptide administered per dose will be in the range of about 0.01μg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1μg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

3.9 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF),

platelet-derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti- inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

3.9.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome

coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

3.9.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol,

propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose,

hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or

aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without

destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T

cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are

useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering

agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final

composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

3.9.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical

procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 μg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

3.9.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

3.10 ANTIBODIES

Another aspect of the invention is an antibody that specifically binds the polypeptide of the invention. Such antibodies include monoclonal and polyclonal antibodies, single chain antibodies, chimeric antibodies, bifunctional/bispecific antibodies, humanized antibodies, human antibodies, and complementary determining region (CDR)-grafted antibodies, including compounds which include CDR and/or antigen-binding sequences, which specifically recognize a polypeptide of the invention. Preferred antibodies of the invention are human antibodies which are produced and identified according to methods described in WO93/11236, published June 20, 1993, which is incorporated herein by reference in its entirety. Antibody fragments, including Fab, Fab', $F(ab')_2$, and F_v, are also provided by the invention. The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor , NY (1988),

Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins. Antibodies of the invention can be produced using any method well known and routinely practiced in the art.

Non-human antibodies may be humanized by any methods known in the art. In one method, the non-human CDRs are inserted into a human antibody or consensus antibody framework sequence. Further changes can then be introduced into the antibody framework to modulate affinity or immunogenicity.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Polypeptides of the invention may also be used to immunize animals to obtain polyclonal and monoclonal antibodies which specifically react with the protein. Such antibodies may be obtained using either the entire protein or fragments thereof as an immunogen. The peptide immunogens additionally may contain a cysteine residue at the carboxyl terminus, and are conjugated to a hapten such as keyhole limpet hemocyanin (KLH). Methods for synthesizing such peptides are known in the art, for example, as in R. P. Merrifield, J. Amer. Chem. Soc. 85, 2149-2154 (1963); J. L. Krstenansky, et al., FEBS Lett. 211, 10 (1987).

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions

associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein. In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A.M., Monoclonal Antibodies Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth et al., J. Immunol. 35:1-21 (1990); Kohler and Milstein, Nature 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., Immunology Today 4:72 (1983); Cole et al., in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985), pp. 77-96).

Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with a peptide or polypeptide of the invention. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection. The protein that is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to, coupling the antigen with a heterologous protein (such as globulin or β-galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, Western blot analysis, or radioimmunoassay (Lutz et al., Exp. Cell Research. 175:109-124 (1988)). Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell,

A.M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)). Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody-containing antiserum is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The present invention further provides the above- described antibodies in delectably labeled form. Antibodies can be delectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.) fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well-known in the art, for example, see (Sternberger, L.A. et al., J. Histochem. Cytochem. 18:315 (1970); Bayer, E.A. et al., Meth. Enzym. 62:308 (1979); Engval, E. et al., Immunol. 109:129 (1972); Goding, J.W. J. Immunol. Meth. 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

3.11 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NOs: 1 - 362 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NOs: 1 - 362 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence

information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based

systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

3.12 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 15241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

3.13 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays:

Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

3.14 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

3.15 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NOs: 1 - 362, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives

expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a

skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

3.16 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NOs: 1 - 362. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from of any of the nucleotide sequences SEQ ID NOs: 1 - 362 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

3.17 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6) 1469-72); using UV light (Nagata *et al.,* 1985; Dahlen *et al.,* 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.,* 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8) 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.,* (1991) Anal. Biochem. 198(1) 138-42).

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., (1991). In this technology, a phosphoramidate bond is employed (Chu et al., (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond

joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-$MeIm_7$), is then added to a final concentration of 10 mM 1-$MeIm_7$. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-$MeIm_7$, is made fresh and 25 ul added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res. 19(12) 3345-50; or linked to Teflon using the method of

Duncan & Cavalier (1988) Anal. Biochem. 169(1) 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.,* (1994) PNAS USA 91(11) 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

3.18 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24) 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to

the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*JI, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*JI normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*JI**), yield a quasi-random distribution of DNA fragments form the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*JI** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *Cvi*JI** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

3.19 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the

density of the wells is achieved. One to 25 dots may be accommodated in 1 mm^2, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

4.0 EXAMPLES

4.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

4.2 EXAMPLE 2

Novel Nucleic Acids

The novel nucleic acids of the present invention of the invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 120, gb pri 120, UniGene version 120, Genepet release 120). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1- 362.

Table 1 shows the various tissue sources of SEQ ID NO: 1-362.

The homology for SEQ ID NO: 1-362 were obtained by a BLASTP version 2.0al 19MP-WashU search against Genpept release 120 and the amino acid version of Geneseq released on October 26, 2000, using BLAST algorithm. The results showed homologues for SEQ ID NO: 1-362 from Genpept. The homologues with identifiable functions for SEQ ID NO: 1-362 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren

Brunak, and Gunnar von Heijne in the publication " Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et as reference, was obtained for the polypeptide sequences. Table 5 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

TABLE 1

TISSUE ORIGIN	LIBRARY/ RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
adult brain	GIBCO	AB3001	4 18 39-40 83 88 98 110 112-113 136 168-169 201-203
adult brain	GIBCO	ABD003	7 15-16 31-32 39-41 45 54 58 63 70 73-75 82-84 92 98 106 110 114 116-117 126 128 130 139 144 155 164 168-169 191-192 195 198 204-215 239-240 249 252 258 272-274
adult brain	Clontech	ABR001	10-11 15 19 39-40 88 106 120 144 168 215-216 258
adult brain	Clontech	ABR006	13 17 20 23 33 39-40 50 58 62 75 82 84 88 100 104 121-122 129 149 168 208 216 223 232-233 239 256 269 277 287-288 353 360
adult brain	Clontech	ABR008	4 10-11 13 17 20 23 25 28-30 32 34-35 39-41 48 50 53-54 58 61 63 68-69 74 76 78 80 84-89 91 98 104 107 112-114 118 121-122 130 134-136 143 153-155 158-160 163-166 168 172-173 184-188 199-200 203 212-213 215-216 219-220 226-227 234 239 242 244 251-252 255-257 263 268 271-272 277-280 287 291 300-301 305-306 316 322 338 346-347 360
adult brain	Clontech	ABR011	157 306
adult brain	BioChain	ABR012	36 247
adult brain	Invitrogen	ABR013	176
adult brain	Invitrogen	ABR014	50 53 100 269
adult brain	Invitrogen	ABR015	19 38 74 161-162
brain	Invitrogen	ABR016	53 74 137 139 239
adult brain	Invitrogen	ABT004	8 15 19-20 28 30 35 75 78 100 106-107 113 134 160 179 181 184 198-199 210 216 224 227 252 254-255 288 340
adipocytes	Stratagene	ADP001	9 13 19 45 74 98 121-122 131 164 187 189-190 217 239
adrenal gland	Clontech	ADR002	9 15 18-19 24-25 31-32 46 56 77-78 112 114-115 117-119 121-122 124 139 170 182 192 209 213 218 220 225 249 276 306
adult heart	GIBCO	AHR001	2 4 7 17 19-22 26-27 34 38 45-46 50 53-54 58 60-61 63 74 76-77 86-87 91 96 98 108 112 114 121-122 131 133 136-140 144 155 160 165-168 184 188 217 226 239 241-242 251 259 265 277-278 290 306
adult kidney	GIBCO	AKD001	4 6-11 13 15-17 19-20 24 30-32

TISSUE ORIGIN	LIBRARY/ RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
			34 36-38 47 53-54 60-63 66 69 73-75 78 82-85 87 89-92 96 98 100 103 106 108 110 112-113 116 121-123 126 129 131 134 136 139-142 144 153 155 158-159 169-170 176 181 207 237 239 266-267 271-272 306
adult kidney	Invitrogen	AKT002	7-8 10-11 13 15 19 25-27 32 37-38 53 55-56 66 75 86 90 92 108 123 144 165-166 172 182 199 218 225 233 236 238 260 266-267 332
adult lung	GIBCO	ALG001	8 22 26-28 38-40 47 54 78 91 98 104 110 112 117 139 148 168 189 196 225 239 248 351-352
lymph node	Clontech	ALN001	7 26-27 32 35 38-40 79 82 120 127 152 158-159 169 171 219 239 244
young liver	GIBCO	ALV001	7 14 16-17 19 33 37 53 72 77 107 113 116 118 134 152 168 212 249
adult liver	Invitrogen	ALV002	12 14 17 24 28 32-33 36 58 73 75-76 84 101 116 131 138 140 158-160 182 194 212 238 275 284 323 342-343
adult liver	Clontech	ALV003	271 284 358
ovary	Invitrogen	AOV001	4 6-11 13 15-16 18-21 25-27 31-32 34 36 38-40 46 48 50 53-54 56 58 60 65 70 73-78 80 83-84 86 91-92 95 98 100-101 103-106 108 110-112 115 117-118 124 126-127 129-131 136 139-142 144 148 155 157-161 163-167 169 173-174 178 180-186 188-189 191-193 196 199-200 204-208 210-211 220-223 233 236 239 249-252 260-263 266-270 287-288 306 315 351-352
placenta	Clontech	APL001	30 50 74 82 230
placenta	Invitrogen	APL002	45 50 59 70 75 103 163 223
adult spleen	GIBCO	ASP001	7 19 30 38 45 54 58 62 74 81 83 91 106 110 112-113 116 131 144 151 155 162 165-166 172 176 189 191 215 230 236 239 249 329
testis	GIBCO	ATS001	4 15 19-20 30 48 53 74 89 94 110 126 140 158-159 173 214-215 220 239 245 306
bladder	Invitrogen	BLD001	30 35 59 61 74-75 123 164 221 241 318
bone marrow	Clontech	BMD001	3 6-7 9 13 17 20 26-27 30-31 34 38-40 42 46 53-54 63-79 82-83 85 91 93-98 101 105 110 115 121-122 126 128-129 133-134

TISSUE ORIGIN	LIBRARY/ RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
			143 145 154 161-162 176 192 205-206 234 236 239 243 264 289 306 322
bone marrow	Clontech	BMD002	3-4 7 9 13 16-17 19-20 23 30 32 34 36 38-40 47-48 54-56 58 61 68-69 74-75 79 84 108 118-119 121-122 125 128-129 131 133 140 144 147 149 153-154 158-159 161 163 167 171 174 176 185-187 200 211 218 232 239 241 247 252 277-278 285 296 303 310 320 324 329 339 341 353 356 359
bone marrow	Clontech	BMD004	64
colon	Invitrogen	CLN001	18 32 100 106 110 143 153 163-164 178 213 247 266-267 284
cervix	BioChain	CVX001	4 6 8-9 19 22 24-25 28 32 45-46 53 55-56 63 74-75 77-78 83 87 91-92 95 102 105 108 110 123 127 136-137 140 169 172 182 184-186 189-191 199 211 238 249 266-267 274 283 306-308 317 354
endothelial cells	Stratagene	EDT001	2 4 6-7 9 15 17-21 25-28 30 32 36 39-40 45 47-48 53 55-57 60 62-63 69-70 74-76 78 83 85-87 98 101-104 106 108 112-113 119 121-123 130-131 136-137 139-142 155-156 158-159 161 174 189-192 204 208 218 220 223 230 239 251 280 306
Genomic clones from short arm of chromosome 8	Genomic Data from Genetic Research	EPM001	223
Genomic clones from short arm of chromosome 8	Genomic Data from Genetic Research	EPM003	223
Genomic clones from short arm of chromosome 8	Genomic Data from Genetic Research	EPM004	223
fetal brain	Clontech	FBR001	32 227
fetal brain	Clontech	FBR004	319
fetal brain	Clontech	FBR006	7 10-11 13 17 20 23-25 28-29 32 35 41-42 48 50 53 63 75 80 89 91 104 112 121-122 125 130 154 163 165-166 168 171 173 191 199 210 215-216 218 226 232 239 256 272 277 290 300 306 309 319-320 333 353 360
fetal brain	Invitrogen	FBT002	15 17 19 35 69 75 87 104 109 140 163 174 192-193 198-199

TISSUE ORIGIN	LIBRARY/ RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
			207 220 228 239 252 256-258
fetal heart	Invitrogen	FHR001	3 8 19 32 41 48 77-79 91 114 119 126 163 165-166 172 174 176 200 218 232 244 263 331 351-352 360-361
fetal kidney	Clontech	FKD001	16-17 36 46 53 74 82 95 104 111 117 169 189
fetal kidney	Clontech	FKD002	26-27 165-166 218 220 232 238 263 306
fetal kidney	Invitrogen	FKD007	38 74
fetal lung	Clontech	FLG001	32 48 139 173 217
fetal lung	Invitrogen	FLG003	10-11 19 36 58 61 69 74 134 163 168 178 194 249 263 266-267 351-352
fetal liver-spleen	Columbia University	FLS001	1-19 21-38 41-62 68 70 72 74-78 87 90-91 93 100-104 106-121 123-125 127 130-131 133-134 141-142 144 149 155-156 161 163 165-167 169 176 194-196 200 207 210 221 224-225 227 231-233 236 238 263 303 306 313 324 336 342
fetal liver-spleen	Columbia University	FLS002	2 5 7-9 12 14 16-18 22-24 30-33 35-40 43-46 48-50 52-53 57 70 72 76-78 84-85 87 90 92 101-102 106-108 110 112 114 116-120 124 127-128 130-131 134-135 140-142 144 155 163 172 174 187 189-190 192 195-196 199 205-207 210 220-221 223-224 230-234 244 251 258 260-261 263 265 275 296 313-315 331 337-338 345 362
fetal liver-spleen	Columbia University	FLS003	19 30 33 139 174 265 313 339 355
fetal liver	Invitrogen	FLV001	10-11 14 17 19 21 37 46 50 61 63 156 163 165-166 172 200 210 238 253
fetal liver	Clontech	FLV002	19 32 74 163 356
fetal liver	Clontech	FLV004	3 14 19 32-33 37 42 47-48 50 58 60 82 85 121-122 129 131 152 171 193 272 353
fetal muscle	Invitrogen	FMS001	28 32 39-40 45 48 50 57 74 107 121-122 131 137 139-140 147 173 204 230 281
fetal muscle	Invitrogen	FMS002	19 23 32 34 55-56 80-81 98 121-122 124 131-132 158-159 199 212 230 280-281 353 357 360
fetal skin	Invitrogen	FSK001	2 4 14-15 17-19 22 41 46 50 53 59 72 75-76 81-82 84 94 103 106 113 128 135 140 144 156 164 167 170 174 188 209-210 220 227 230 238-239 254 306

TISSUE ORIGIN	LIBRARY/ RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
			321-322 333-335
fetal skin	Invitrogen	FSK002	4 34 47 54 79 84 113 126-127 129 134 156 192-193 208 223 230 241 277 285 333
fetal spleen	BioChain	FSP001	32 104
umbilical cord	BioChain	FUC001	4 19 22-23 32 38-40 46 55-56 58 61 73-75 91 98-99 103 106 110 112 116 120 123 129-130 139 160 165-166 175 182 230 234 249 251 302
fetal brain	GIBCO	HFB001	6 9 16 19-20 25 32 35-36 39-41 45 48 53-54 56 60 73 80-81 83-92 98 107 112 114 157-159 163 165-166 172 191 197-198 211 226-227 239 350
infant brain	Columbia University	IB2002	6-8 13 15-17 19 21 32 35 41-42 48 50 60-61 77 81 84-85 88 92 104-106 112-113 116 119 134 139 144 160 165-166 168-169 173 176 191 196 199-201 215 223 225 227-228 239 261 285 290 329 339-340 348
infant brain	Columbia University	IB2003	7-9 13 32 39-41 58 92 103 105-106 144 160 162 199 205-206 219 227-228 271 357
infant brain	Columbia University	IBM002	32 88 340
infant brain	Columbia University	IBS001	6 26-27 32 164 199 340
lung, fibroblast	Stratagene	LFB001	2 4 18-19 25 39-40 46 53 55-56 106 112 124 129 136 139 146 150 164 169 189-190 215 230 239 260 349
adult lung	Invitrogen	LGT002	2 6 8-11 15-16 19 26-28 30 32 39-40 46 48 50 53-56 60-61 66 72 74-75 85 87 92 94 96 98 103-104 108 110 112-113 117 119-120 124 130-131 139-140 149 152-153 155 158-159 167 169 174 176 178 184 189-190 195-196 217 220 229-230 234-239 248-250 263 265-267 280 286 310 329-330 351-352
lymphocyte	ATCC	LPC001	7 13 16 19 32 39-40 54 63 74 82 96 113 120 126 130-131 133 144 150 178 184-186 223 239 241 260 262 294 305 339
leukocytes	GIBCO	LUC001	1 3-4 7-9 13 16-20 26-27 30 32 34-35 38-40 46 48 51 53-56 63 66 70 72-76 78 82 84-85 87 89 91-92 95-96 101 106 108 110-112 114 116 120-122 126-127 129-133 136 139 144 146-152 164 175-179 187 192 232 236

TISSUE ORIGIN	LIBRARY/ RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
			239 241 266-267 292-294 306 325-327 329 339 359
leukocytes	Clontech	LUC003	7-8 17 55-56 76 84 112 129 131 161-162 176 180 185-186 329
melanoma	Clontech	MEL004	4 13 17 28 30-31 39-40 83 85 92 113 126 129 139 160 162 182 198 232 239 303 324
mammary gland	Invitrogen	MMG001	8-11 16-21 28 30 32 35 41 45 58-59 61 72 74-75 78 84 87 92 103-104 106-107 110 113 115-116 123 128 131 134-135 144 152 163 176 181 183 210 212 220-221 230 234 236 238-239 248 251 260 272-273 275-276 306 331 351-352 360
neuron	Stratagene	NTD001	18-19 39-40 45 74 78 85 91
neuron	Stratagene	NTR001	19 21 57 246 265
neuronal cells	Stratagene	NTU001	8-9 18-19 21 32 81 85 87 128 164 174 184
pituitary gland	Clontech	PIT004	13 47 82 87 98 112 288 354
placenta	Clontech	PLA003	13 48 50 58 77 100 106 112 126 129 152 178 232
prostate	Clontech	PRT001	16 19 22 26-27 32 34 46-47 76-77 92 98 106 112 124 172 214 239 260 280 294
rectum	Invitrogen	REC001	8 10-11 18 30 54 74-76 106 113 123-124 143 163 172 213 220 232 237 260 322-323 340
salivary gland	Clontech	SAL001	8 19 36 74 83 104 118 124 150 176 260 295 304
skin fibroblast	ATCC	SFB002	239
small intestine	Clontech	SIN001	9 17 19 22 32 34 54 57 59-60 73 75 84-85 96 99 107 113 118 134 139 144 149 151 185-187 189 197 199 217 219 221 230-231 248 250 253-254 260 266-267 295 304 356
skeletal muscle	Clontech	SKM001	17 19 39-40 48 89 104 116 131 281
spinal cord	Clontech	SPC001	8 19 32 34 38-40 47 58 61 74 80 83-84 89 104 108 131 139-140 168 187 213 226 236 239 300 350
adult spleen	Clontech	SPLc01	1 46 54 134 236
stomach	Clontech	STO001	7 32 38-40 51 66 74 76 89 117 124 128 169 229 239 253 280 294 296
thalamus	Clontech	THA002	24 30 50 87 124 127 143 163 201 207 220 223 230 266-267 269 279
thymus	Clontech	THM001	7 13 19 25 32 36 39-40 54-56 72 74 82 96 108 113 119 127

TISSUE ORIGIN	LIBRARY/ RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
			137 139 141-142 146 169 184 192 260 276 296
thymus	Clontech	THMc02	9 17 28 30 32 39-40 48 53 61 72 74-75 77 79 82 91 107 112 119-122 125-126 131 139-142 153 171 175-176 178 184 187 205-206 222-223 227 235-236 269 278 289 297 305 310-311 325 327-329 336
thyroid gland	Clontech	THR001	7-11 15 17 19-20 25-27 32 34 36 46 48 53 59 72 82-87 89 91 96 98-99 104 106 110 118-119 121-122 127 130 136 139 144 151-152 158-159 165-167 179 187 204 208 220 239 249 281 283 295 298-299 312 316 344
trachea	Clontech	TRC001	62-63 73 75 86-87 89 101 147 192 239 266-267 282-283
uterus	Clontech	UTR001	4 8 17 19 22 26-27 32 39-40 46 63 82 98 110 130 151

TABLE 2

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
1	L29075	Dictyostelium discoideum G-box binding factor	173	21
2	AL359215	Streptomyces coelicolor A3(2) putative phosphoglycerate mutase.	133	28
3	AF228713	Homo sapiens EDAG-1	1671	100
4	AC007130	Homo sapiens similar to 3-hydroxyisobutyrate dehydrogenase ; similar to P29266 (PID:g416873)	1557	100
5	AB040926	Homo sapiens KIAA1493 protein	1973	98
6	AF193016	Homo sapiens methyltransferase COQ3	1609	99
7	U95825	Homo sapiens androgen-induced prostate proliferative shutoff associated protein	2968	63
8	AL390081	Homo sapiens SEMA4B, Semaphorin 4B	3560	99
9	AC002130	Arabidopsis thaliana F1N21.9	258	50
10	Z38061	Saccharomyces cerevisiae mal5, sta1, len: 1367, CAI: 0.3, AMYH_YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	323	27
11	D88733	Equine herpesvirus 1 membrane glycoprotein	284	24
13	M80783	Homo sapiens B12 protein	1144	70
14	U72678	Mus musculus EF-9	792	92
15	AK026486	Homo sapiens unnamed protein product	427	83
16	AK025813	Homo sapiens unnamed protein product	1010	100
17	AF151036	Homo sapiens HSPC202	722	84
18	AY007148	Homo sapiens similar to Homo sapiens HSPC197 mRNA with GenBank Accession Number AF151031.1	984	100
19	X57432	Rattus rattus ribosomal protein S2	956	97
20	AF164793	Homo sapiens protein x 013	386	100
21	J02642	Homo sapiens glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12)	1639	95
22	M34573	Homo sapiens alpha-2 collagen type VI-a	515	100
23	AL109928	Homo sapiens dJ551D2.5 (novel protein)	1999	100
24	AF111858	Homo sapiens dimethylglycine dehydrogenase precursor	3918	99
25	U64854	Caenorhabditis elegans partial CDS	184	25
26	AF151072	Homo sapiens HSPC238	838	99
27	AF151072	Homo sapiens HSPC238	393	96
28	AK024825	Homo sapiens unnamed protein product	1794	99
29	AF285631	Rattus norvegicus secretory carrier membrane protein 4	894	75
30	AK024113	Homo sapiens unnamed protein	3672	99

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
		product		
31	AL161515	Arabidopsis thaliana putative protein	146	52
32	AJ007798	Homo sapiens stromal antigen 3, (STAG3)	6320	99
33	D31856	Bacillus subtilis HutI protein, imidazolone-5-propionate hydrolase	667	39
34	AL391145	Arabidopsis thaliana putative protein	423	24
35	AF134726	Homo sapiens G7A	1591	46
36	AJ276485	Homo sapiens integral membrane transporter protein	1502	100
37	J05158	Homo sapiens carboxypeptidase N (EC 3.4.17.3)	2274	88
38	X57351	Homo sapiens 1-8D	673	97
39	AF230904	Homo sapiens c-Cbl-interacting protein	3437	100
40	AF230904	Homo sapiens c-Cbl-interacting protein	2615	99
41	AF276893	Homo sapiens p21-activated protein kinase 6	3550	100
42	AF269255	Homo sapiens lysosomal apyrase-like protein 1	3198	100
43	S85655	Homo sapiens prohibitin	742	84
44	AB040926	Homo sapiens KIAA1493 protein	1973	98
45	AF151063	Homo sapiens HSPC229	1012	100
46	X68277	Homo sapiens protein-tyrosine phosphatase	1886	100
47	Z98745	Homo sapiens dJ29K1.2	889	51
48	AF032668	Rattus norvegicus rsec15	3738	92
50	AF195534	Rattus norvegicus GERp95	4513	99
51	AF161368	Homo sapiens HSPC105	513	98
52	W73147	Amino acid sequence of the soluble complement receptor 1	651	81
53	AF271212	Homo sapiens disrupter of silencing SAS10	2431	100
54	AF116646	Homo sapiens PRO0082	598	100
55	AF145613	Drosophila melanogaster BcDNA.GH03108	817	46
56	AF145613	Drosophila melanogaster BcDNA.GH03108	884	38
57	AL023803	Homo sapiens dJ616B8.3 (novel gene)	2287	100
59	AC024877	Caenorhabditis elegans contains similarity to Pfam families PF00621 (Guanine nucleotide exchange factor for Rho/Rac/Cdc42-like GTPases, score=58.2, E=1.7e-13, N=10 and PF00169 (PH (pleckstrin homology) domain, score=17.0, E=0.00071, N=1)	296	31
60	AL390114	Leishmania major probable proteophosphoglycan	154	30
61	AL031427	Homo sapiens dJ167A19.1 (novel	732	51

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
		protein)		
62	AL390935	Leishmania major possible CG17807 protein	151	43
63	J04067	Canis familiaris microsomal signal peptidase	930	99
64	AF062378	Mus musculus calmodulin-binding protein SHA1	1782	60
65	AE001002	Archaeoglobus fulgidus ATP-dependent RNA helicase, putative	195	29
66	X69065	Erythroid ankyrin [Mus musculus]	181	30
68	AF017807	Homo sapiens Arp2/3 complex 16kDa subunit	371	100
69	AC007660	Arabidopsis thaliana putative translation initiation factor	173	29
70	AJ243177	Xenopus laevis Xenopus RPA interacting protein alpha	447	42
71	AF226055	Homo sapiens HTGN29	1367	100
72	AF090930	Homo sapiens PRO0478	180	89
73	AF118084	Homo sapiens PRO1914	350	98
74	AB028893	Homo sapiens ribosomal protein S11	824	100
75	AK024500	Homo sapiens FLJ00109 protein	1514	100
76	AF238866	Mus musculus LNR42	1041	99
77	AC026875	Arabidopsis thaliana T6D22.6	129	30
78	U42436	Caenorhabditis elegans Hypothetical protein C49H3.3	130	32
79	M80902	Homo sapiens AHNAK nucleoprotein	8529	99
80	W90962	Human CSGP-2 protein [homo sapiens]	2346	99
81	AF206661	Gallus gallus neuronal tetraspanin	1066	81
82	S73591	Homo sapiens brain-expressed HHCPA78 homolog VDUP1	800	42
83	AF116650	Homo sapiens PRO0786	302	100
84	L26335	Cavia porcellus zinc finger protein	1493	99
85	AF209198	Homo sapiens zinc finger protein 277	2357	100
86	AE001399	Plasmodium falciparum GAF domain protein (cyclic nt signal transduct.)	178	35
87	Y48226	Human prostate cancer-associated protein 12 [Homo sapiens]	1204	96
88	M94389	Loligo pealei neurofilament protein	165	23
89	AF121775	Homo sapiens nasopharyngeal carcinoma susceptibility protein LZ16	903	58
90	AF116675	Homo sapiens PRO1942	257	100
91	AE002760	Drosophila melanogaster CG14464 gene product	195	43
92	AK000100	Homo sapiens unnamed protein product	841	100
93	AB020236	Homo sapiens ribosomal protein L27A	754	99
94	AF119865	Homo sapiens PRO2176	470	97
96	AF138863	Homo sapiens PRO1677	868	99
97	X14361	Homo sapiens CR-1 receptor SCR9 (or	135	100

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
		16) C-term. (21 is 3rd base in codon) (106 is 1st base in codon)		
98	Z24725	Homo sapiens mitogen inducible gene mig-2	3576	99
99	U64598	Caenorhabditis elegans weakly similar to S. cervisiae PTM1 precursor (SP:P32857)	398	45
100	AC004770	Homo sapiens BC269730_4	1527	84
101	AL139075	Campylobacter jejuni NOL1\NOP2\sun family protein	312	35
102	AF113694	Homo sapiens PRO1359	416	100
103	U15158	Homo sapiens ESP-2	564	41
104	AL020996	Homo sapiens dJ317E23.3 (novel protein)	1818	99
105	AF161370	Homo sapiens HSPC107	824	100
106	AK000161	Homo sapiens unnamed protein product	284	100
107	AK001784	Homo sapiens unnamed protein product	684	100
108	AE000913	Methanobacterium thermoautotrophicum conserved protein	221	25
109	AF165527	Homo sapiens DGCR8	859	100
110	AF230200	Homo sapiens OVN6-2	358	95
111	Z72516	Caenorhabditis elegans T25G3.1	180	36
112	AF201940	Homo sapiens DC6	505	100
113	AK001301	Homo sapiens unnamed protein product	2040	98
114	U23515	Caenorhabditis elegans weakly similar to gastrula zinc finger protein	205	47
115	AF228021	Bos taurus cyclophilin I	345	91
116	AF166124	Homo sapiens selenoprotein X	527	100
117	AF079445	Dictyostelium discoideum TipC	529	30
118	AB032179	Homo sapiens similar to mouse Ehm2	2255	100
119	U89867	Homo sapiens nuclear matrix protein 55	2449	98
120	U29056	Mus musculus Src-like adapter protein	352	47
121	U22015	Mus musculus retinoid X receptor interacting protein	2190	73
122	AF113538	Homo sapiens retinoid x receptor interacting protein	1800	100
123	AK000158	Homo sapiens unnamed protein product	740	100
124	AF260924	Mus musculus UFD2/D4COLE1E fusion protein	1222	82
125	U12465	Homo sapiens ribosomal protein L35	591	97
126	AJ277591	Homo sapiens p15-2a protein	749	100
127	AF205599	Mus musculus transposase-like protein	2406	74
128	U58975	Homo sapiens proto-oncogene	659	90

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
129	X98374	Rattus norvegicus KIS	2193	99
130	AF151049	Homo sapiens HSPC215	627	100
131	M59807	Homo sapiens putative	907	99
132	U12979	Homo sapiens PC4	563	99
133	AF076642	Homo sapiens regulator of G-protein signaling 13	1218	100
134	AF116718	Homo sapiens PRO2900	396	100
135	AC018758	Homo sapiens GPI-anchored metastasis-associated protein homolog	213	31
136	AC025416	Arabidopsis thaliana F5O11.12	135	36
137	M83186	Homo sapiens cytochrome c oxidase subunit VIIa	247	100
138	AF232937	Mus musculus thymic stroma derived lymphopoietin	247	41
139	M15841	Homo sapiens U2 small nuclear ribonucleoprotein B''	638	100
140	AK026916	Homo sapiens unnamed protein product	2612	99
141	Y05317	Human secreted protein bn97_1 [Homo sapiens]	1508	100
142	Y05317	Human secreted protein bn97_1 [Homo sapiens]	851	99
143	AF041083	Rattus norvegicus RoBo-1	139	25
144	AC024260	Arabidopsis thaliana cell division control protein, putative; 15914-18846	194	25
146	AL022398	Homo sapiens dJ434O14.3.2 (putative protein) (isoform 2)	575	100
147	AF212842	Homo sapiens immunoglobulin-like transcript 11 protein	1280	99
148	AB042827	Rattus norvegicus Nadrin	477	66
149	AK001841	Homo sapiens unnamed protein product	1916	100
150	AJ278120	Homo sapiens putative ankyrin-repeat containing protein	540	98
151	AL135959	Homo sapiens dJ233G16.1 (novel protein)	770	100
152	Y58196	[Homo sapiens] Human STRAP-3 protein, encoded by testis EST AI139607	671	100
153	U41060	Homo sapiens LIV-1 protein	373	50
154	AJ007590	Homo sapiens XRP2 protein	1766	100
155	AB046868	Xenopus laevis beta-catenin-interacting protein	125	46
156	AB027258	Homo sapiens basal transcriptional activator hABT1	1408	100
157	AF039656	Homo sapiens neuronal tissue-enriched acidic protein	1109	96
158	AK001425	Homo sapiens unnamed protein product	1695	99
159	AK001425	Homo sapiens unnamed protein	858	98

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
		product		
160	AK002030	Homo sapiens unnamed protein product	1029	100
161	X79417	Sus scrofa 40S ribosomal protein S12	510	83
162	X12597	Homo sapiens HMG-1 protein (AA 1-215)	1140	99
163	AK001159	Homo sapiens unnamed protein product	764	100
164	AK000020	Homo sapiens unnamed protein product	1613	100
165	AK001322	Homo sapiens unnamed protein product	1207	100
166	AK001322	Homo sapiens unnamed protein product	892	98
167	AE003822	Drosophila melanogaster CG8493 gene product	357	36
168	AF023451	Bos taurus guanine nucleotide-exchange protein	187	21
169	AK000154	Homo sapiens unnamed protein product	673	100
170	AJ132702	Mus musculus ATFa-associated factor	435	64
172	AL022311	Homo sapiens dJ1014D13.3 (novel protein)	405	38
174	AB017634	Mus musculus ENP	770	65
175	U40407	synthetic construct T cell receptor alpha chain	1119	80
176	AF043179	Homo sapiens T cell receptor beta chain	681	73
177	AF116678	Homo sapiens PRO1995	587	100
178	AF217522	Homo sapiens uncharacterized bone marrow protein BM046	262	42
179	AB046074	Macaca fascicularis unnamed protein product	515	83
180	X79417	Sus scrofa 40S ribosomal protein S12	429	84
181	AF002668	Homo sapiens MLD	1235	65
182	AB036422	Bos taurus molybdopterin cofactor sulfurase	3509	79
184	AF036696	Caenorhabditis elegans contains similarity to Brassica oleracea non-green plastid phosphate/triose-phosphate translocator precursor (GB:U13632)	662	42
185	AJ277276	Homo sapiens rapa-2	5155	99
186	AJ277275	Homo sapiens rapa-1	5086	100
187	U22296	Rattus norvegicus casein kinase 1 gamma 1 isoform	1444	93
188	AE003750	Drosophila melanogaster CG9996 gene product	468	44
189	Z97056	Homo sapiens dJ434P1.2 (KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum	1103	100

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
		protein retention receptor 3)		
190	AF081126	Drosophila melanogaster ER lumen protein retaining receptor	409	75
192	AF226047	Homo sapiens GL002	863	100
193	AF269167	Homo sapiens arsenite related gene 1	906	60
195	U41805	Mus musculus putative T1/ST2 receptor binding protein precursor	162	26
197	AL357374	Homo sapiens bA353C18.2 (novel protein)	404	97
199	M34551	Homo sapiens 52-kD Ro/SSA ribonucleoprotein	964	42
202	AF230201	Homo sapiens OVC10-2	396	100
203	AK001984	Homo sapiens unnamed protein product	658	100
204	AK000530	Homo sapiens unnamed protein product	691	100
205	U37134	Drosophila melanogaster inturned protein	248	23
206	U37134	Drosophila melanogaster inturned protein	244	23
208	AB033130	Mus musculus testis-specific gene	871	85
209	AK000464	Homo sapiens unnamed protein product	221	100
210	AJ277557	Homo sapiens mitochondrial 5'(3')-deoxyribonucleotidase (dNT-2)	617	100
211	AF127564	Arabidopsis thaliana ubiquitin-protein ligase 1	854	42
213	Y17108	Homo sapiens rhomboid-related protein	485	39
214	AL132776	Homo sapiens dJ393D12.2 (novel LIM domain protein)	1660	99
215	U73819	Mus musculus polypeptide GalNAc transferase-T4	1039	42
216	AL035406	Homo sapiens dJ233K16.1 (KIAA0444, a putative chromodomain helicase DNA binding protein 3 (CHD3))	3844	100
217	M15800	Homo sapiens MAL protein	308	42
218	L29554	Rattus norvegicus alpha 2,6-sialyltransferase	942	80
219	AL137315	Homo sapiens hypothetical protein	983	100
220	AK026027	Homo sapiens unnamed protein product	647	100
221	AL137584	Homo sapiens hypothetical protein	246	97
223	AC005498	Homo sapiens R31665_1	1752	78
225	AC010155	Arabidopsis thaliana F3M18.5	171	34
226	AL080276	Homo sapiens dJ101K10.2 (regulator of G-protein signaling 17 (RGS17) (RGSZ2))	1126	100
227	AF042345	Homo sapiens truncated EVI5	1815	64
228	J04214	Bos taurus retinaldehyde-binding protein precursor	504	39

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
230	AF181263	Homo sapiens EH domain containing 2	2816	99
231	AP001660	Homo sapiens putative gene, multidrug resistance associated protein like	1424	100
232	AB000910	Sus scrofa ribosomal protein	542	100
233	AL133404	Homo sapiens dJ238O23.9 (novel protein similar to rat SAC (soluble adenylyl cyclase))	298	100
234	X51397	Mus musculus MyD88 protein (AA 1-243)	136	25
235	X01403	Homo sapiens T-cell receptor alpha-chain	840	90
236	X14254	Rattus rattus invariant chain (AA 1-280)	745	77
238	U23084	Saccharomyces cerevisiae Ynl0470p	344	35
239	X03342	Homo sapiens rpL32 (aa 1-135)	152	96
240	AF116669	Homo sapiens PRO1828	237	100
241	U23181	Caenorhabditis elegans final exon in repeat region; similar to long tandem repeat region of sialidase (SP:TCNA_TRYCR, P23253) and neurofilament H protein	135	25
242	AF263913	Mus musculus fidgetin	3864	97
243	AF090892	Homo sapiens PRO0106	290	100
244	U21310	Caenorhabditis elegans F40H6.3 gene product	153	27
246	AK001673	Homo sapiens unnamed protein product	3661	100
247	AL022603	Arabidopsis thaliana putative protein	166	43
248	AL023803	Homo sapiens dJ616B8.3 (novel gene)	339	42
249	X52140	Rattus norvegicus precursor polypeptide (AA -28 to 1152)	5429	87
250	AB020755	Arabidopsis thaliana gene_id:MZN1.18~unknown protein	139	46
251	AE003619	Drosophila melanogaster CG7224 gene product	186	43
252	AC004997	Homo sapiens match to ESTs Z43979 (NID:g573097), R19699 (NID:g774333), and C01164 (NID:g1433394); alternatively spliced form of H_DJ130H16.1a (C-terminal truncation confirmed by C01164)	388	67
254	AE003588	Drosophila melanogaster CG13947 gene product	115	42
256	Y50934	Human fetal brain cDNA clone vc30_1 derived protein #1 [Homo sapiens]	498	100
257	AF242768	Homo sapiens mesenchymal stem cell protein DSC43	1554	100
259	M95779	Bos taurus G protein gamma-5 subunit	333	98

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
260	AL035521	Arabidopsis thaliana putative protein	145	28
261	AF247501	Drosophila melanogaster PINEAPPLE EYE	333	36
263	AL034548	Homo sapiens dJ1103G7.2 (novel protein)	262	100
264	AF119851	Homo sapiens PRO1722	143	63
265	L41834	Ensis minor nuclear protein	173	26
266	X97966	Homo sapiens calcyphosine	963	100
267	X97966	Homo sapiens calcyphosine	660	95
269	AF022383	Homo sapiens complexin I	668	99
271	Y10054	Rattus norvegicus 3-hydroxy-3-methylglutaryl CoA lyase	224	67
274	AF153201	Homo sapiens zinc finger protein dp	179	36
275	X85738	Bos taurus novel brain-specific protein	326	55
277	AF250342	Arabidopsis thaliana SMC-related protein MSS2	266	39
278	AL080242	Homo sapiens bA554C12.1 (RBX1 or ROC1 (ring-box or ring finger protein 1))	131	100
279	Z83760	Ciona intestinalis COS41.4	1162	62
280	U41534	Caenorhabditis elegans similar to yeast MAK16 protein (SP:MK16_YEAST,P10962)	721	54
281	AF272975	Gallus gallus smoothelin-C	543	37
282	AL035414	Homo sapiens dJ667H12.2.2 (novel protein (isoform 2))	588	100
283	AF116661	Homo sapiens PRO1438	145	62
285	AK001757	Homo sapiens unnamed protein product	1300	100
287	U20897	Homo sapiens melanoma ubiquitous mutated protein	2133	100
289	U09847	Homo sapiens zinc finger protein	880	100
290	AJ000079	Trypanosoma cruzi glycosylphosphatidylinositol-specific phospholipase C	225	26
291	AF156549	Mus musculus putative E1-E2 ATPase	2108	49
293	AF161345	Homo sapiens HSPC082	439	100
294	AF116694	Homo sapiens PRO2219	351	88
295	M74027	Homo sapiens mucin	461	39
298	AL133640	Homo sapiens hypothetical protein	2149	100
299	M17886	Homo sapiens acidic ribosomal phosphoprotein (P1)	161	76
300	Y99368	Human PRO1326 (UNQ686) amino acid sequence SEQ ID NO:100 [Homo sapiens]	300	32
303	AE003708	Drosophila melanogaster CG6171 gene product	144	27
304	M32639	Homo sapiens statherin precursor	276	87
305	Z83844	Homo sapiens dJ37E16.2 (SH3-domain binding protein 1)	897	96

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
306	AE003791	Drosophila melanogaster CG18065 gene product	120	32
307	AF135026	Homo sapiens kallikrein-like protein 3 splice variant 1	1392	100
310	AF198257	Felis catus immunoglobulin kappa light chain	678	76
311	X57725	Homo sapiens TCR Vbeta 22a	626	100
312	AC018513	Homo sapiens unknown	818	100
313	X03249	Bos taurus epsilon-4 beta-globin	321	79
314	AB046099	Macaca fascicularis unnamed protein product	395	88
315	AC006033	Homo sapiens T cell receptor gamma chain; match to S08328 (PID:g106470)	1017	95
316	AB046103	Macaca fascicularis unnamed protein product	801	94
317	U88895	Homo sapiens ORF2	399	81
318	U09848	Homo sapiens zinc finger protein	242	49
319	AB003184	Homo sapiens ISLR	880	59
320	AB036921	Chrysophrys major maturation-inducing protein	797	69
322	AF284422	Homo sapiens cation-chloride cotransporter-interacting protein	4694	100
325	AE000659	Homo sapiens TCRAV8S2	577	100
327	R59748	T cell receptor Valpha2.3 chain [homo sapiens]	636	100
328	AJ004871	Homo sapiens TCR alpha chain	1328	94
329	AF043179	Homo sapiens T cell receptor beta chain	1286	92
330	AF090930	Homo sapiens PRO0478	140	50
332	AF077043	Homo sapiens 60S ribosomal protein L36	275	87
333	AL121988	Homo sapiens dJ34M23.3 (gap junction protein, beta 4 (connexin 30.3))	1457	100
334	D86424	Mus musculus high-sulfur keratin protein	521	87
335	AF090434	Fundulus heteroclitus cytochrome P450 2N1	760	40
336	AF116688	Homo sapiens PRO2133	370	98
337	X85372	Homo sapiens Sm protein F	222	84
338	D87009	Homo sapiens putative	1822	99
339	AE000860	Methanobacterium thermoautotrophicum conserved protein	631	35
340	AL049759	Homo sapiens dJ930L11.1 (similar to KIAA0397)	1305	98
341	AE000004	Mycoplasma pneumonia MG207 homolog, from M. genitalium	141	27
342	AF151076	Homo sapiens HSPC242	135	100
343	AB037902	Homo sapiens truncated aldo-keto reductase	670	100

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
345	M33014	Drosophila melanogaster ubiquitin	153	62
346	AF053356	Homo sapiens leucin rich neuronal protein	580	46
348	AL137512	Homo sapiens hypothetical protein	751	100
349	S68015	Homo sapiens c6.1A	1664	100
350	AF151037	Homo sapiens HSPC203	318	100
351	AB036432	Homo sapiens advanced glycation endproducts receptor	2133	100
352	AB036432	Homo sapiens advanced glycation endproducts receptor	2094	96
353	AC006942	Homo sapiens R31181_2, partial protein	547	100
354	AF125535	Homo sapiens pp21 homolog	502	95
355	AF227130	Homo sapiens candidate taste receptor T2R3	1629	100
357	AB046626	Macaca fascicularis hypothetical protein	291	93
358	Z69597	Canis familiaris Rod transducin alpha subunit	1145	100
359	AE000659	Homo sapiens TCRAV16S1	565	100
360	Y99368	Human PRO1326 (UNQ686) Amino acid sequence SEQ ID NO:100. [Homo sapiens]	2034	100
362	L06499	Homo sapiens ribosomal protein L37a	187	55

TABLE 3

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
1	PR00651	5-HYDROXYTRYPTAMINE 2B RECEPTOR SIGNATURE	PR00651E 10.53 4.025e-06 60-80
2	BL00126	3'5'-cyclic nucleotide phosphodiesterases proteins.	BL00126B 15.20 7.750e-06 208-220
3	DM00892	3 RETROVIRAL PROTEINASE.	DM00892C 23.55 9.438e-07 285-319
4	BL00895	3-hydroxyisobutyrate dehydrogenase proteins.	BL00895B 21.14 7.061e-22 151-190 BL00895C 20.10 8.071e-22 200-236 BL00895A 12.61 1.973e-18 42-63
5	DM00099	4 kw A55R REDUCTASE TERMINAL DIHYDROPTERIDINE.	DM00099A 5.17 5.263e-06 409-415
7	PF00598	Influenza Matrix protein (M1).	PF00598C 19.35 3.333e-07 531-563
9	DM00522	499 kw TRYPSIN KINASE KUNITZ PANCREATIC.	DM00522A 8.30 3.250e-06 287-297
10	PR00514	5-HYDROXYTRYPTAMINE 1D RECEPTOR SIGNATURE	PR00514C 11.01 9.061e-07 81-100
11	PR00514	5-HYDROXYTRYPTAMINE 1D RECEPTOR SIGNATURE	PR00514C 11.01 9.061e-07 81-100
12	PR00775	90 KD HEAT SHOCK PROTEIN SIGNATURE	PR00775G 10.64 3.487e-07 8-27
13	PR00902	VP6 BLUE-TONGUE VIRUS INNER CAPSID PROTEIN SIGNATURE	PR00902K 11.09 1.000e-05 176-200
14	PR00875	MOLLUSC METALLOTHIONEIN SIGNATURE	PR00875A 5.83 1.127e-07 159-171 PR00875D 5.00 1.000e-05 158-169
15	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803C 7.00 9.200e-07 181-191 DM01803A 10.51 1.000e-06 178-199 DM01803C 7.00 7.337e-06 214-224
16	PF00803	3A movement protein.	PF00803D 14.15 2.622e-06 41-71
17	PR00170	SODIUM CHANNEL SIGNATURE	PR00170G 7.74 1.000e-05 24-53
18	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701E 13.83 5.684e-06 117-133
20	DM01415	6 SALIVARY GLUE PROTEIN.	DM01415A 6.65 6.063e-06 55-68
21	DM01418	352 FIBRILLAR COLLAGEN CARBOXYL-TERMINAL.	DM01418A 20.83 7.731e-06 64-112
22	BL00616	Histidine acid phosphatases phosphohistidine proteins.	BL00616D 15.83 7.268e-06 117-133

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
23	DM00611	9 kw LECTIN HTPG SERINE GNTR.	DM00611A 7.73 5.826e-06 173-181
24	BL00832	2'-5'-oligoadenylate synthetases proteins.	BL00832D 21.81 5.017e-06 425-449
25	PR00354	7FE FERREDOXIN SIGNATURE	PR00354C 5.72 8.590e-09 543-561
26	PR00217	43 KD POSTSYNAPTIC PROTEIN SIGNATURE	PR00217C 10.91 3.851e-07 89-105
27	PR00513	5-HYDROXYTRYPTAMINE 1B RECEPTOR SIGNATURE	PR00513A 7.75 1.439e-06 168-180
28	DM00552	GROWTH FACTOR AND CYTOKINES RECEPTORS FAMILY.	DM00552A 11.97 1.000e-05 130-152
29	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701I 8.59 3.088e-06 102-126
30	DM00060	338 kw NEUREXIN ALPHA III CYSTEINE.	DM00060 6.92 3.284e-07 680-690
34	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517D 7.66 6.971e-07 484-496
35	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701A 14.28 4.183e-06 722-744
37	PR00513	5-HYDROXYTRYPTAMINE 1B RECEPTOR SIGNATURE	PR00513C 10.79 8.927e-07 287-304
38	PR00166	AROMATIC AMINO ACID PERMEASE SIGNATURE	PR00166I 11.06 1.000e-05 98-118
39	PR00003	4-DISULPHIDE CORE SIGNATURE	PR00003A 14.69 3.803e-06 311-321
40	PR00003	4-DISULPHIDE CORE SIGNATURE	PR00003A 14.69 3.803e-06 311-321
41	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517D 7.66 8.788e-07 2-14
42	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701F 14.45 7.750e-06 25-46
43	DM00895	7 kw REVERSE TRANSCRIPTASE RNA POLYMERASE.	DM00895B 8.85 4.185e-06 157-167
44	DM00099	4 kw A55R REDUCTASE TERMINAL DIHYDROPTERIDINE.	DM00099A 5.17 5.263e-06 409-415
45	PR00519	5-HYDROXYTRYPTAMINE 5B RECEPTOR SIGNATURE	PR00519A 8.06 8.984e-06 137-154
46	PR00519	5-HYDROXYTRYPTAMINE 5B RECEPTOR SIGNATURE	PR00519B 9.99 1.828e-07 151-168
47	DM00892	3 RETROVIRAL PROTEINASE.	DM00892B 9.78 2.047e-06 21-27
48	BL00832	2'-5'-oligoadenylate synthetases	BL00832B 15.45 6.836e-07 375-416

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
		proteins.	
49	DM00588	8 kw CHO2 ALPHA ANTIGEN PARAMYOSIN.	DM00588A 10.87 7.128e-06 20-31
50	DM00604	2 SHIGA/RICIN RIBOSOMAL INACTIVATING TOXINS.	DM00604D 13.26 8.250e-06 263-273
51	BL01193	Ribosomal protein S8e proteins.	BL01193A 13.21 1.000e-05 19-50
52	PR00172	GLUCOSE TRANSPORTER SIGNATURE	PR00172F 8.47 9.901e-06 69-90
53	PR00297	10 KD CHAPERONIN SIGNATURE	PR00297A 13.91 4.740e-06 379-395
54	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320A 16.74 9.710e-06 81-96
55	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517C 5.36 4.126e-07 352-365
56	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517C 5.36 4.126e-07 352-365
57	DM00547	1 kw CHROMO BROMODOMAIN SHADOW GLOBAL.	DM00547E 13.94 4.656e-06 229-252
58	PF00506	Influenza virus nucleoprotein.	PF00506I 10.26 3.723e-06 32-68
60	DM00522	499 kw TRYPSIN KINASE KUNITZ PANCREATIC.	DM00522B 9.43 7.338e-07 171-185
61	DM01123	5 kw RESISTANCE TETRACYCLINE METHYLENOMYCIN EXPORT.	DM01123B 20.06 3.187e-06 205-244
62	PR00439	11-S SEED STORAGE PROTEIN FAMILY SIGNATURE	PR00439G 17.85 9.239e-07 82-100
63	PR00652	5-HYDROXYTRYPTAMINE 7 RECEPTOR SIGNATURE	PR00652F 11.66 4.767e-06 100-122
64	DM01785	72 PYRUVATE (FLAVODOXIN) DEHYDROGENASE.	DM01785A 14.90 2.196e-06 218-261
65	PR00652	5-HYDROXYTRYPTAMINE 7 RECEPTOR SIGNATURE	PR00652A 8.92 5.104e-06 315-336
67	BL00405	43 Kd postsynaptic protein.	BL00405F 8.07 9.920e-06 13-44
68	PR00380	KINESIN HEAVY CHAIN SIGNATURE	PR00380D 9.93 9.043e-06 44-66
69	PR00683	SPECTRIN PLECKSTRIN HOMOLOGY DOMAIN SIGNATURE	PR00683D 15.87 9.571e-06 46-65
70	PR00753	1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE SIGNATURE	PR00753C 13.93 7.330e-06 192-213
71	PF00602	Influenza RNA-dependant RNA	PF00602J 9.52 9.727e-06 47-102

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
		polymerase subunit PB1.	
72	DM01855	PROTEIN-GLUTAMATE O-METHYLTRANSFERASE.	DM01855A 11.54 7.594e-06 27-44
74	BL01277	Ribonuclease PH proteins.	BL01277A 17.39 1.000e-05 50-88
75	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517G 16.45 6.919e-06 755-771
77	PF01073	3-beta hydroxysteriod dehydrogenase/isomerase family.	PF01073B 12.26 9.767e-07 102-147
78	PR00351	MAS20 PROTEIN IMPORT RECEPTOR SIGNATURE	PR00351C 7.03 6.182e-06 99-112 PR00351C 7.03 1.000e-05 5-18
79	DM00611	9 kw LECTIN HTPG SERINE GNTR.	DM00611C 11.08 4.549e-06 1489-1501
80	DM01111	4 kw PHOSPHATASE TRANSFORMING 61K PDF1.	DM01111C 9.35 2.800e-06 44-73
82	PD02407	3-BISPHOSPHOGLYCERATE-INDEPENDENT PHOSPHOGLYCER.	PD02407B 16.51 1.000e-06 94-111
83	PR00116	ARGINASE SIGNATURE	PR00116D 14.91 9.850e-06 14-44
84	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517F 11.48 7.250e-06 45-62
87	DM00522	499 kw TRYPSIN KINASE KUNITZ PANCREATIC.	DM00522B 9.43 3.535e-07 223-237
88	DM00303	6 LEA 11-MER REPEAT REPEAT.	DM00303A 13.20 8.034e-08 270-320
91	DM01242	3 THREONINE--TRNA LIGASE.	DM01242B 23.57 4.672e-06 71-120
92	PR00388	3',5'-CYCLIC NUCLEOTIDE CLASS II PHOSPHODIESTERASE SIGNATURE	PR00388E 6.66 3.797e-06 124-136
93	PD02407	3-BISPHOSPHOGLYCERATE-INDEPENDENT PHOSPHOGLYCER.	PD02407B 16.51 9.676e-06 15-32
94	PF00506	Influenza virus nucleoprotein.	PF00506I 10.26 4.555e-06 16-52
95	PR00551	2-S GLOBULIN FAMILY SIGNATURE	PR00551H 11.29 8.740e-06 21-39
96	PR00756	MEMBRANE ALANYL DIPEPTIDASE (M1) FAMILY SIGNATURE	PR00756E 11.91 9.338e-06 68-81
97	PR00547	X OPIOID RECEPTOR SIGNATURE	PR00547B 6.96 3.268e-06 17-36
98	PR00651	5-HYDROXYTRYPTAMINE 2B RECEPTOR	PR00651A 16.53 4.000e-06 653-674

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
		SIGNATURE	
99	PR00208	GLIADIN AND LMW GLUTENIN SUPERFAMILY SIGNATURE	PR00208C 11.51 9.775e-06 54-71
101	PR00451	CHITIN-BINDING DOMAIN SIGNATURE	PR00451A 6.49 1.000e-05 152-161
102	BL00832	2'-5'-oligoadenylate synthetases proteins.	BL00832B 15.45 7.569e-06 1-42
103	BL00405	43 Kd postsynaptic protein.	BL00405J 13.28 6.952e-06 142-176
104	DM01242	3 THREONINE--TRNA LIGASE.	DM01242F 10.61 5.500e-07 187-201
105	DM01834	8 HYDROGENASE (FE) SMALL CHAIN.	DM01834A 4.96 7.097e-06 53-60
107	PR00101	ASPARTATE CARBAMOYLTRANSFERASE SIGNATURE	PR00101E 5.52 1.000e-05 111-117
109	PR00902	VP6 BLUE-TONGUE VIRUS INNER CAPSID PROTEIN SIGNATURE	PR00902K 11.09 9.922e-06 91-115
110	PR00259	TRANSMEMBRANE FOUR FAMILY SIGNATURE	PR00259A 9.27 9.716e-06 9-33
111	BL00785	5'-nucleotidase proteins.	BL00785B 10.65 6.507e-06 53-67
112	PR00652	5-HYDROXYTRYPTAMINE 7 RECEPTOR SIGNATURE	PR00652G 10.94 5.429e-06 14-32
113	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517D 7.66 3.661e-06 372-384
114	PF00637	7-fold repeat proteins in Clathrin, also in VPS proteins.	PF00637D 7.09 9.449e-07 32-44
115	DM01235	5 kw T4 55.10 METHYLCYTOSINE TRANSCRIPTASE.	DM01235 20.29 9.832e-06 77-108
116	PR00873	ECHINOIDEA (SEA URCHIN) METALLOTHIONEIN SIGNATURE	PR00873C 6.16 9.906e-06 70-81
117	PR00387	3'5'-CYCLIC NUCLEOTIDE PHOSPHODIESTERASE SIGNATURE	PR00387D 10.81 4.889e-06 155-172
118	PF00598	Influenza Matrix protein (M1).	PF00598A 14.24 7.158e-06 211-254
119	BL00895	3-hydroxyisobutyrate dehydrogenase proteins.	BL00895B 21.14 8.036e-06 428-467
120	PR00419	ADRENODOXIN REDUCTASE FAMILY SIGNATURE	PR00419D 10.62 9.430e-06 18-33
121	DM00396	5 kw INTRON COI ND4L	DM00396B 7.85 3.739e-07 381-389

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
		ND5.	
122	BL00198	4Fe-4S ferredoxins, iron-sulfur binding region proteins.	BL00198 10.43 5.500e-06 135-147
123	DM01554	1 THYROLIBERIN PRECURSOR.	DM01554E 10.78 4.208e-06 93-110
124	BL00405	43 Kd postsynaptic protein.	BL00405G 7.78 6.294e-06 130-167
125	PR00298	60 KD CHAPERONIN SIGNATURE	PR00298D 10.23 3.847e-06 14-40
126	PR00317	EPENDYMIN SIGNATURE	PR00317A 13.39 9.897e-06 79-99
127	PR00513	5-HYDROXYTRYPTAMINE 1B RECEPTOR SIGNATURE	PR00513B 17.51 3.971e-06 277-290
130	PR00516	5-HYDROXYTRYPTAMINE 2A RECEPTOR SIGNATURE	PR00516G 15.11 8.811e-06 18-35
131	PR00828	FORMIN SIGNATURE	PR00828F 8.56 1.000e-05 61-81
132	DM01269	303 kw ACTIVATING RAN GTPASE ISOZYME.	DM01269A 23.35 7.279e-06 28-56
133	PR00586	PROSTANOID EP4 RECEPTOR SIGNATURE	PR00586B 14.97 7.322e-06 10-28 PR00586H 8.65 9.791e-06 16-40
134	PF00954	S-locus glycoprotein family.	PF00954D 18.68 9.843e-06 9-44
135	PR00018	KRINGLE DOMAIN SIGNATURE	PR00018A 14.52 1.000e-05 120-136
136	BL00115	Eukaryotic RNA polymerase II heptapeptide repeat proteins.	BL00115E 14.13 9.921e-06 40-69
137	PR00521	ANDROGEN RECEPTOR SIGNATURE	PR00521A 17.02 9.729e-06 5-25
138	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701I 8.59 5.267e-07 16-40
139	DM01269	303 kw ACTIVATING RAN GTPASE ISOZYME.	DM01269A 23.35 7.085e-08 93-121
140	PR00915	LUTEOVIRUS GROUP 1 COAT PROTEIN SIGNATURE	PR00915D 16.14 1.000e-05 374-392
143	DM00522	499 kw TRYPSIN KINASE KUNITZ PANCREATIC.	DM00522A 8.30 4.441e-06 94-104
144	PF00598	Influenza Matrix protein (M1).	PF00598B 13.10 1.623e-06 89-133
145	PR00217	43 KD POSTSYNAPTIC PROTEIN SIGNATURE	PR00217C 10.91 6.250e-06 24-40
147	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701A 14.28 6.049e-06 266-288
148	PR00513	5-HYDROXYTRYPTAMINE 1B RECEPTOR SIGNATURE	PR00513D 11.06 9.920e-06 103-121
149	PF00603	Influenza RNA-dependant RNA polymerase subunit	PF00603D 8.49 9.319e-07 30-85

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
		PA.	
150	DM01688	2 POLY-IG RECEPTOR.	DM01688N 11.93 9.920e-08 72-100
151	PF00637	7-fold repeat proteins in Clathrin, also in VPS proteins.	PF00637B 10.68 6.906e-06 186-195
152	BL00461	6-phosphogluconate dehydrogenase proteins.	BL00461A 15.90 1.764e-08 21-57
153	DM01554	1 THYROLIBERIN PRECURSOR.	DM01554A 6.07 2.565e-06 589-599
154	BL00405	43 Kd postsynaptic protein.	BL00405E 8.84 8.125e-06 109-135
155	PF00637	7-fold repeat proteins in Clathrin, also in VPS proteins.	PF00637A 15.49 5.179e-06 42-65
156	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517D 7.66 7.339e-06 85-97
157	DM01688	2 POLY-IG RECEPTOR.	DM01688P 13.54 1.925e-07 44-89 DM01688L 4.36 2.367e-07 123-133
159	PR00933	B-LYTIC METALLOENDOPEPTIDASE (M23) SIGNATURE	PR00933D 13.92 1.000e-05 85-106
161	PR00352	3FE-4S FERREDOXIN SIGNATURE	PR00352A 11.15 6.162e-06 94-106
163	BL00785	5'-nucleotidase proteins.	BL00785E 15.85 4.000e-06 95-111
164	PR00916	2C ENDOPEPTIDASE (C24) CYSTEINE PROTEASE FAMILY SIGNATURE	PR00916C 8.02 2.655e-06 121-133
165	BL00785	5'-nucleotidase proteins.	BL00785D 9.89 3.045e-06 154-164
166	BL00785	5'-nucleotidase proteins.	BL00785D 9.89 3.045e-06 123-133
167	DM01023	2 GLYCOSYL HYDROLASES FAMILY 5.	DM01023C 13.51 6.486e-06 149-175
168	PF00803	3A movement protein.	PF00803A 15.38 8.088e-06 255-290
169	PR00282	SNAKE CYTOTOXIN SIGNATURE	PR00282D 11.82 9.882e-06 74-85
170	PR00519	5-HYDROXYTRYPTAMINE 5B RECEPTOR SIGNATURE	PR00519B 9.99 5.787e-06 83-100
172	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803A 10.51 6.804e-06 136-157
173	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701E 13.83 9.724e-06 14-30
174	BL00118	Phospholipase A2 histidine proteins.	BL00118A 16.00 9.842e-06 132-145
175	DM01688	2 POLY-IG RECEPTOR.	DM01688G 16.45 1.825e-06 89-121
176	DM01930	2 kw FINGER SMCX SMCY YDR096W.	DM01930A 7.97 2.403e-07 146-159

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
177	PR00510	NEBULIN SIGNATURE	PR00510F 9.88 8.552e-06 34-51
179	PF00432	Prenyltransferase and squalene oxidase repeat proteins.	PF00432A 11.90 1.000e-05 27-39
180	PR00537	MU OPIOID RECEPTOR SIGNATURE	PR00537A 8.17 1.000e-05 27-41
183	PR00536	MELANOCYTE STIMULATING HORMONE RECEPTOR SIGNATURE	PR00536C 8.58 8.833e-06 64-82
184	DM00973	3 kw RESISTANCE BENOMYL YLL028W CYCLOHEXIMIDE.	DM00973B 17.81 8.261e-06 158-184
185	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517C 5.36 4.265e-06 718-731
186	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517C 5.36 4.265e-06 718-731
187	PR00651	5-HYDROXYTRYPTAMINE 2B RECEPTOR SIGNATURE	PR00651A 16.53 6.447e-06 144-165
188	BL01017	Ergosterol biosynthesis ERG4/ERG24 family proteins.	BL01017D 20.82 9.737e-06 21-67
191	DM00315	072 RIBONUCLEASE INHIBITOR.	DM00315B 6.84 7.459e-06 95-107
192	PR00930	HIGH MOBILITY GROUP PROTEIN (HMGY) SIGNATURE	PR00930E 5.98 9.740e-06 285-298
193	BL00794	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase proteins.	BL00794B 22.12 8.967e-06 150-191
194	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517C 5.36 5.853e-06 115-128
196	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803A 10.51 7.031e-07 11-32
197	PR00171	SUGAR TRANSPORTER SIGNATURE	PR00171B 14.73 1.000e-05 15-35
198	PD02407	3-BISPHOSPHOGLYCERATE-INDEPENDENT PHOSPHOGLYCER.	PD02407J 10.55 6.610e-06 69-81
199	PF00604	Influenza RNA-dependant RNA polymerase subunit PB2.	PF00604F 10.21 2.417e-06 276-331
201	PR00409	PHTHALATE DIOXYGENASE REDUCTASE FAMILY SIGNATURE	PR00409D 13.02 9.900e-06 43-58
202	BL00660	Band 4.1 family	BL00660A 31.50 9.595e-06 1-54

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
		domain proteins.	
203	PR00745	GLYCOSYL HYDROLASE FAMILY 39 SIGNATURE	PR00745D 15.85 9.700e-06 68-83
204	PD01364	MUCIN GLYCOPROTEIN PRECURSOR MEM.	PD01364A 6.18 9.667e-06 9-16
205	BL00794	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase proteins.	BL00794C 19.62 7.690e-07 309-347
206	BL00794	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase proteins.	BL00794C 19.62 7.690e-07 309-347
207	DM00895	7 kw REVERSE TRANSCRIPTASE RNA POLYMERASE.	DM00895E 15.72 3.170e-06 241-266
208	PR00517	5-HYDROXYTRYPTAMINE 2C RECEPTOR SIGNATURE	PR00517D 7.66 4.835e-06 59-71
209	PD02365	CHAIN FACTOR INTERLEUKIN-12 BETA PRECURSOR IL-1.	PD02365C 7.89 9.719e-06 20-50
210	PR00551	2-S GLOBULIN FAMILY SIGNATURE	PR00551E 10.27 9.432e-06 19-34
211	DM01418	352 FIBRILLAR COLLAGEN CARBOXYL-TERMINAL.	DM01418B 22.51 3.289e-06 527-569
213	DM01785	72 PYRUVATE (FLAVODOXIN) DEHYDROGENASE.	DM01785E 12.98 6.400e-06 165-216
214	DM01834	8 HYDROGENASE (FE) SMALL CHAIN.	DM01834B 15.29 3.382e-06 64-90
215	PR00217	43 KD POSTSYNAPTIC PROTEIN SIGNATURE	PR00217C 10.91 4.583e-06 407-423
216	DM00547	1 kw CHROMO BROMODOMAIN SHADOW GLOBAL.	DM00547F 23.43 6.538e-36 628-675 DM00547E 13.94 2.400e-18 387-410 DM00547C 17.30 9.486e-16 266-288 DM00547B 11.28 9.217e-15 237-251 DM00547D 11.60 4.951e-12 357-371 DM00547A 12.38 6.455e-11 216-228
217	BL00407	Connexins proteins.	BL00407D 17.61 1.000e-05 57-87
218	BL00198	4Fe-4S ferredoxins, iron-sulfur binding region proteins.	BL00198 10.43 9.481e-06 74-86
219	DM01417	6 kw INDUCING XPMC2 MUSHROOM SPAC22G7.04.	DM01417B 15.47 3.550e-06 90-102
220	PR00519	5-HYDROXYTRYPTAMINE 5B RECEPTOR SIGNATURE	PR00519E 3.58 2.404e-07 184-199
221	PD01313	INTRON PROBABLE MATURASE CHLOROPLAST MR.	PD01313B 23.27 1.000e-05 10-45
222	PR00047	C4-TYPE STEROID	PR00047A 15.70 9.878e-06 99-116

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
		RECEPTOR ZINC FINGER SIGNATURE	
223	DM01554	1 THYROLIBERIN PRECURSOR.	DM01554C 11.76 3.571e-06 255-271
224	DM01123	5 kw RESISTANCE TETRACYCLINE METHYLENOMYCIN EXPORT.	DM01123B 20.06 5.206e-06 28-67
226	BL00198	4Fe-4S ferredoxins, iron-sulfur binding region proteins.	BL00198 10.43 8.630e-07 28-40
227	BL00895	3-hydroxyisobutyrate dehydrogenase proteins.	BL00895B 21.14 5.173e-06 185-224
229	PR00907	THROMBOMODULIN SIGNATURE	PR00907G 11.63 9.794e-06 13-40
230	PR00651	5-HYDROXYTRYPTAMINE 2B RECEPTOR SIGNATURE	PR00651B 9.95 6.416e-06 62-77
231	DM00611	9 kw LECTIN HTPG SERINE GNTR.	DM00611C 11.08 9.113e-06 214-226
232	PR00582	PROSTANOID EP3 RECEPTOR SIGNATURE	PR00582B 9.74 1.000e-05 76-95
233	PR00407	EUKARYOTIC MOLYBDOPTERIN DOMAIN SIGNATURE	PR00407E 13.51 9.438e-06 176-192
234	DM00892	3 RETROVIRAL PROTEINASE.	DM00892C 23.55 9.913e-06 6-40
235	DM01688	2 POLY-IG RECEPTOR.	DM01688G 16.45 4.450e-06 94-126 DM01688J 14.69 6.000e-06 34-71
236	PD00930	PROTEIN GTPASE DOMAIN ACTIVATION.	PD00930A 25.62 1.000e-05 80-106
237	PR00076	6-PHOSPHOGLUCONATE DEHYDROGENASE SIGNATURE	PR00076B 11.24 6.418e-07 14-44
239	PR00243	MUSCARINIC ACETYLCHOLINE RECEPTOR SIGNATURE	PR00243F 16.45 9.182e-06 7-18
240	BL00854	Proteasome B-type subunits proteins.	BL00854B 10.97 1.000e-05 1-9
241	PR00513	5-HYDROXYTRYPTAMINE 1B RECEPTOR SIGNATURE	PR00513B 17.51 5.263e-07 876-889
242	DM01111	4 kw PHOSPHATASE TRANSFORMING 61K PDF1.	DM01111I 15.32 2.473e-07 522-552
243	BL00415	Synapsins proteins.	BL00415B 9.91 9.778e-06 53-89
244	BL00514	Fibrinogen beta and gamma chains C-terminal domain proteins.	BL00514E 14.28 1.000e-05 221-238
245	PR00187	ARTHROPOD HAEMOCYANIN SIGNATURE	PR00187B 15.70 1.000e-05 37-55

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
246	PF00637	7-fold repeat proteins in Clathrin, also in VPS proteins.	PF00637C 27.33 1.184e-06 368-415
247	BL00785	5'-nucleotidase proteins.	BL00785A 9.73 7.557e-06 57-68
248	PR00651	5-HYDROXYTRYPTAMINE 2B RECEPTOR SIGNATURE	PR00651D 12.56 2.615e-06 228-249
249	PR00516	5-HYDROXYTRYPTAMINE 2A RECEPTOR SIGNATURE	PR00516B 10.78 1.811e-06 310-325
250	BL00461	6-phosphogluconate dehydrogenase proteins.	BL00461C 18.34 9.495e-06 30-58
251	BL00888	Cyclic nucleotide-binding domain proteins.	BL00888A 18.03 9.667e-06 20-37
252	DM01242	3 THREONINE--TRNA LIGASE.	DM01242E 23.00 6.215e-07 119-161
253	DM00250	kw ANNEXIN ANTIGEN PROLINE TUMOR.	DM00250A 10.52 6.488e-06 16-32
254	BL00291	Prion protein.	BL00291A 4.49 2.469e-07 51-86 BL00291A 4.49 6.878e-07 40-75 BL00291A 4.49 5.330e-06 22-57 BL00291A 4.49 1.000e-05 30-65
255	PF00506	Influenza virus nucleoprotein.	PF00506F 9.40 5.459e-08 17-55
256	BL00126	3'5'-cyclic nucleotide phosphodiesterases proteins.	BL00126A 27.56 6.026e-06 25-62
257	PR00003	4-DISULPHIDE CORE SIGNATURE	PR00003D 8.10 5.131e-06 291-300
258	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803C 7.00 7.061e-06 10-20
259	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803C 7.00 8.071e-06 3-13
260	PR00775	90 KD HEAT SHOCK PROTEIN SIGNATURE	PR00775D 8.91 3.831e-06 147-165
261	PR00217	43 KD POSTSYNAPTIC PROTEIN SIGNATURE	PR00217C 10.91 1.167e-06 75-91
262	PR00388	3',5'-CYCLIC NUCLEOTIDE CLASS II PHOSPHODIESTERASE SIGNATURE	PR00388D 14.87 8.079e-06 69-83
263	PR00023	ZONA PELLUCIDA SPERM-BINDING PROTEIN SIGNATURE	PR00023A 17.17 9.036e-06 24-39
264	BL00024	Hemopexin domain proteins.	BL00024F 11.30 9.894e-06 3-24
265	DM00303	6 LEA 11-MER REPEAT REPEAT.	DM00303A 13.20 6.294e-06 177-227
266	PR00652	5-HYDROXYTRYPTAMINE	PR00652A 8.92 9.224e-07 69-90

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
		7 RECEPTOR SIGNATURE	
267	PR00652	5-HYDROXYTRYPTAMINE 7 RECEPTOR SIGNATURE	PR00652A 8.92 9.224e-07 69-90
268	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803A 10.51 1.673e-06 14-35
270	PR00875	MOLLUSC METALLOTHIONEIN SIGNATURE	PR00875C 8.64 9.550e-06 65-77
271	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803A 10.51 7.308e-06 21-42
272	PF00598	Influenza Matrix protein (M1).	PF00598A 14.24 4.383e-06 46-89
273	PR00113	ALKALINE PHOSPHATASE SIGNATURE	PR00113D 6.87 9.260e-06 8-19
274	PD01841	PHOSPHORYLASE KINASE ALPHA MUSCL.	PD01841E 18.60 9.446e-06 80-118
275	PF00600	Influenza non-structural protein (NS1).	PF00600A 20.40 1.563e-06 40-67
276	PR00877	PLANT PEC FAMILY METALLOTHIONEIN SIGNATURE	PR00877B 4.74 9.878e-06 31-38
277	PR00076	6-PHOSPHOGLUCONATE DEHYDROGENASE SIGNATURE	PR00076E 12.73 6.417e-06 71-99
279	BL00101	Hexapeptide-repeat containing-transferases proteins.	BL00101A 10.95 1.000e-05 71-78
280	DM01111	4 kw PHOSPHATASE TRANSFORMING 61K PDF1.	DM01111M 10.67 2.629e-06 163-187
282	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 9.934e-06 35-50
283	PR00519	5-HYDROXYTRYPTAMINE 5B RECEPTOR SIGNATURE	PR00519C 9.73 1.227e-06 22-37
284	PR00304	TAILLESS COMPLEX POLYPEPTIDE 1 (CHAPERONE) SIGNATURE	PR00304E 7.79 1.000e-05 54-67
285	BL00197	2Fe-2S ferredoxins, iron-sulfur binding region proteins.	BL00197A 18.23 9.866e-07 49-79
286	PR00753	1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE SIGNATURE	PR00753D 6.85 8.636e-06 61-83
287	PR00003	4-DISULPHIDE CORE SIGNATURE	PR00003B 7.64 1.300e-06 166-174
288	BL00940	Gamma-thionins family proteins.	BL00940A 20.51 9.671e-06 16-40
289	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 9.609e-11 122-135 PD00066 13.92 1.900e-09 94-107 PD00066 13.92 2.703e-07 66-79

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
			PD00066 13.92 1.000e-05 38-51
291	PR00516	5-HYDROXYTRYPTAMINE 2A RECEPTOR SIGNATURE	PR00516F 10.18 9.609e-07 761-779
293	PR00651	5-HYDROXYTRYPTAMINE 2B RECEPTOR SIGNATURE	PR00651E 10.53 8.487e-06 51-71
296	PR00635	AT1 ANGIOTENSIN II RECEPTOR SIGNATURE	PR00635C 7.44 8.602e-06 28-45
297	PF00915	Calicivirus coat protein.	PF00915E 5.71 1.000e-05 102-112
298	PR00519	5-HYDROXYTRYPTAMINE 5B RECEPTOR SIGNATURE	PR00519B 9.99 3.968e-06 495-512
299	PR00784	MITOCHONDRIAL BROWN FAT UNCOUPLING PROTEIN SIGNATURE	PR00784D 15.86 9.730e-06 22-40
300	DM00973	3 kw RESISTANCE BENOMYL YLL028W CYCLOHEXIMIDE.	DM00973A 21.17 1.273e-06 7-44
301	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 8.752e-06 42-57
302	BL00283	Soybean trypsin inhibitor (Kunitz) protease inhibitors family.	BL00283B 16.55 1.000e-05 15-30
303	DM01753	6 kw OSTEOBLAST MAJOR IMMUNOGENIC MPB70.	DM01753A 21.93 9.830e-06 59-94
304	PR00516	5-HYDROXYTRYPTAMINE 2A RECEPTOR SIGNATURE	PR00516E 14.87 6.516e-06 20-38
305	DM01554	1 THYROLIBERIN PRECURSOR.	DM01554E 10.78 5.452e-07 20-37
306	PR00331	HAEMAGGLUTININ HA2 CHAIN SIGNATURE	PR00331E 18.67 1.000e-05 75-93
307	PR00651	5-HYDROXYTRYPTAMINE 2B RECEPTOR SIGNATURE	PR00651H 5.59 8.858e-06 152-175
308	BL00208	Plant hemoglobins proteins.	BL00208A 18.41 1.000e-05 5-47
309	PR00240	ALPHA-1A ADRENERGIC RECEPTOR SIGNATURE	PR00240E 9.25 9.391e-06 36-56
310	PR00701	60KD INNER MEMBRANE PROTEIN SIGNATURE	PR00701C 10.53 8.255e-06 55-76
311	PR00423	CELL DIVISION PROTEIN FTSZ SIGNATURE	PR00423B 7.15 1.000e-05 5-26
312	PF00602	Influenza RNA-dependant RNA polymerase subunit PB1.	PF00602C 12.16 2.068e-07 26-66
313	PR00246	SOMATOSTATIN RECEPTOR SIGNATURE	PR00246D 7.36 1.000e-05 4-14

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
314	BL00216	Sugar transport proteins.	BL00216A 13.29 9.526e-06 26-38
316	PF00721	Tobacco mosaic virus coat.	PF00721A 14.59 9.845e-06 131-167
317	PD00489	PROTEIN TRANSMEMBRANE TRANSPORT C.	PD00489A 15.57 1.000e-05 55-71
318	PR00163	RUBREDOXIN SIGNATURE	PR00163A 10.47 9.888e-06 59-76
319	PR00513	5-HYDROXYTRYPTAMINE 1B RECEPTOR SIGNATURE	PR00513A 7.75 9.149e-07 205-217
320	DM01418	352 FIBRILLAR COLLAGEN CARBOXYL-TERMINAL.	DM01418C 20.48 5.142e-06 377-419
321	BL00067	3-hydroxyacyl-CoA dehydrogenase proteins.	BL00067D 21.49 7.441e-06 9-42
322	DM01857	5 kw NUCLEOSIDE TRANSPORT DEPENDENT NA.	DM01857B 14.94 7.821e-08 52-80
323	DM01803	1 HERPESVIRUS GLYCOPROTEIN H.	DM01803C 7.00 5.133e-06 43-53
324	PD01672	+ TRANSPORT EXCHANGER NA H TRANS.	PD01672B 15.16 1.000e-05 6-55
325	DM01688	2 POLY-IG RECEPTOR.	DM01688J 14.69 5.538e-06 31-68
327	DM00372	CARCINOEMBRYONIC ANTIGEN PRECURSOR AMINO-TERMINAL DOMAIN.	DM00372A 19.18 1.000e-05 9-54
328	DM01688	2 POLY-IG RECEPTOR.	DM01688J 14.69 4.308e-06 31-68
329	DM01930	2 kw FINGER SMCX SMCY YDR096W.	DM01930A 7.97 2.403e-07 144-157
330	PF00685	Sulfotransferase proteins.	PF00685A 19.12 9.370e-06 49-82
331	PR00347	PATHOGENESIS-RELATED PROTEIN SIGNATURE	PR00347A 13.98 9.649e-06 55-68
332	PR00538	MUSCARINIC M1 RECEPTOR SIGNATURE	PR00538F 10.59 8.667e-06 30-48
334	PR00159	2FE-2S FERREDOXIN SIGNATURE	PR00159A 9.58 1.153e-06 23-32
336	PR00554	ADENOSINE A2B RECEPTOR SIGNATURE	PR00554B 12.52 9.778e-06 41-50
337	DM01111	4 kw PHOSPHATASE TRANSFORMING 61K PDF1.	DM01111G 10.39 7.250e-06 3-44
338	PR00516	5-HYDROXYTRYPTAMINE 2A RECEPTOR SIGNATURE	PR00516B 10.78 7.649e-06 214-229
340	PR00388	3',5'-CYCLIC NUCLEOTIDE CLASS II PHOSPHODIESTERASE SIGNATURE	PR00388A 10.45 5.050e-06 141-160
342	DM01664	kw.	DM01664D 16.63 1.000e-05 22-47

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
343	PD01841	PHOSPHORYLASE KINASE ALPHA MUSCL.	PD01841K 14.81 1.000e-05 65-95
344	PR00416	EUKARYOTIC DNA TOPOISOMERASE I SIGNATURE	PR00416D 12.12 9.772e-06 23-40
345	BL00726	AP endonucleases family 1 proteins.	BL00726C 19.90 1.000e-05 7-33
346	BL00305	11-S plant seed storage proteins.	BL00305D 21.08 4.566e-06 465-507
347	PR00332	HISTIDINE TRIAD FAMILY SIGNATURE	PR00332A 10.15 9.890e-06 16-33
348	PR00518	5-HYDROXYTRYPTAMINE 5A RECEPTOR SIGNATURE	PR00518A 8.62 7.807e-06 19-36
349	BL00305	11-S plant seed storage proteins.	BL00305D 21.08 4.736e-06 276-318
350	PR00503	BROMODOMAIN SIGNATURE	PR00503C 19.84 9.731e-06 28-47
351	DM01688	2 POLY-IG RECEPTOR.	DM01688K 17.19 9.066e-07 81-120
352	DM01688	2 POLY-IG RECEPTOR.	DM01688K 17.19 9.066e-07 81-120
353	DM01415	6 SALIVARY GLUE PROTEIN.	DM01415B 13.78 5.273e-06 99-147
354	PR00216	OSTEOPONTIN SIGNATURE	PR00216F 11.79 9.913e-06 50-69
356	DM00895	7 kw REVERSE TRANSCRIPTASE RNA POLYMERASE.	DM00895G 3.62 9.913e-06 62-72
357	BL00126	3'5'-cyclic nucleotide phosphodiesterases proteins.	BL00126B 15.20 6.329e-06 35-47
358	PR00512	5-HYDROXYTRYPTAMINE 1A RECEPTOR SIGNATURE	PR00512G 6.54 3.139e-06 3-19
359	PD02455	ELEMENT TRANSPOSABLE INSERTION PROTEIN TRANSPOSITION DNA.	PD02455D 18.65 1.000e-05 58-77
360	DM01415	6 SALIVARY GLUE PROTEIN.	DM01415A 6.65 3.250e-07 16-29
361	BL00794	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase proteins.	BL00794C 19.62 9.702e-06 27-65
362	PR00866	RNA-DEPENDENT DNA-POLYMERASE (MSDNA) SIGNATURE	PR00866B 9.86 9.786e-06 60-73

* Results include in order: accession number subtype; raw score; p-value; postion of signature in amino acid sequence.

TABLE 4

SEQ ID NO:	pFAM NAME	DESCRIPTION	p-value	pFAM SCORE
2	PGAM	Phosphoglycerate mutase family	2.5e-05	23.4
6	Ubie_methyltran	ubiE/COQ5 methyltransferase family	0.035	-133.9
8	Plexin_repeat	Plexin repeat	0.03	18.4
13	K_tetra	K+ channel tetramerisation domain	2.3e-31	117.6
14	EGF	EGF-like domain	7.8e-14	59.4
16	Armadillo_seg	Armadillo/beta-catenin-like repeats	1.3e-05	32.1
19	Ribosomal_S5	Ribosomal protein S5	1.7e-46	167.9
21	gpdh	glyceraldehyde 3-phosphate dehydrogenases	1.3e-230	773.2
24	GCV_T	Glycine cleavage T-protein (aminomethyl tran	9.3e-156	530.9
25	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.015	12.5
26	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	1.6e-10	38.4
33	urease	Urease	0.014	11.0
35	tRNA-synt_1e	tRNA synthetases class I (C)	0.0091	12.1
37	LRRNT	Leucine rich repeat N-terminal domain	0.00049	26.8
39	SH3	SH3 domain	3.4e-60	213.4
40	SH3	SH3 domain	3.4e-60	213.4
41	PBD	P21-Rho-binding domain	1e-08	42.4
42	GDA1_CD39	GDA1/CD39 (nucleoside phosphatase) family	9e-94	324.9
43	Band_7	SPFH domain / Band 7 family	1.7e-21	84.9
46	Rhodanese	Rhodanese-like domain	2.9e-24	94.0
47	zf-C2H2	Zinc finger, C2H2 type	6.2e-32	119.5
50	ZAP	ZAP domain	1.6e-50	181.3
52	sushi	Sushi domain (SCR repeat)	9.5e-27	102.3
55	zf-C2H2	Zinc finger, C2H2 type	0.047	20.3
56	zf-C2H2	Zinc finger, C2H2 type	0.00021	28.1
59	PH	PH domain	2.6e-06	27.6
60	PHD	PHD-finger	2e-09	44.8
64	IQ	IQ calmodulin-binding motif	6.4e-42	152.7
66	ank	Ank repeat	2.7e-23	90.8
69	eIF-1a	Eukaryotic initiation factor 1A	0.0047	-2.4
74	Ribosomal_S17	Ribosomal protein S17	6e-43	148.6
75	LIM	LIM domain containing proteins	0.00067	19.0
80	Phosphodiest	Type I phosphodiesterase / nucleotide py	2.7e-49	177.2
81	transmembrane4	Transmembrane 4 family	6.6e-61	197.7
84	zf-C2H2	Zinc finger, C2H2 type	1.6e-64	227.8
85	zf-C2H2	Zinc finger, C2H2 type	1.4e-07	38.6
89	ank	Ank repeat	4e-31	116.8
93	L15	Ribosomal protein L15	3.5e-21	61.9

SEQ ID NO:	pFAM NAME	DESCRIPTION	p-value	pFAM SCORE
98	Band_41	FERM domain (Band 4.1 family)	0.00015	16.7
101	Nol1_Nop2_Sun	NOL1/NOP2/sun family	4.5e-19	68.6
103	LIM	LIM domain containing proteins	1.3e-30	113.2
113	WD40	WD domain, G-beta repeat	0.00018	28.3
115	pro_isomerase	Cyclophilin type peptidyl-prolyl cis-tr	5.3e-34	120.4
116	DUF25	Domain of unknown function DUF25	1.1e-11	46.9
118	Band_41	FERM domain (Band 4.1 family)	3.2e-77	242.4
119	rrm	RNA recognition motif. (a.k.a. RRM, RBD, or	1.1e-33	125.4
120	SH3	SH3 domain	3e-05	30.9
125	Ribosomal_L29	Ribosomal L29 protein	1.6e-15	65.0
126	NTF2	Nuclear transport factor 2 (NTF2) domain	7.6e-06	32.2
129	rrm	RNA recognition motif. (a.k.a. RRM, RBD, or	0.0016	25.2
130	Fork_head	Fork head domain	1e-28	108.8
132	PC4	Transcriptional Coactivator p15 (PC4)	2.1e-38	141.0
133	RGS	Regulator of G protein signaling domain	2.6e-45	164.0
137	COX7a	Cytochrome c oxidase subunit VIIa	2.3e-40	147.5
139	rrm	RNA recognition motif. (a.k.a. RRM, RBD, or	3.2e-15	64.0
141	lectin_c	Lectin C-type domain	5.1e-05	30.0
142	lectin_c	Lectin C-type domain	5.1e-05	30.0
147	ig	Immunoglobulin domain	9.1e-07	26.9
150	ank	Ank repeat	8.6e-09	42.6
161	Ribosomal_L7Ae	Ribosomal protein L7Ae	0.03	0.8
162	HMG_box	HMG (high mobility group) box	8e-53	188.9
163	PH	PH domain	3e-13	52.4
168	Peptidase_C6	Helper component proteinase	0.0056	7.9
175	ig	Immunoglobulin domain	2.3e-09	35.2
176	ig	Immunoglobulin domain	9.2e-09	33.3
178	WW	WW domain	0.054	17.2
180	Ribosomal_S12e	Ribosomal protein S12e	1.9e-38	141.1
185	myb_DNA-binding	Myb-like DNA-binding domain	0.00011	29.1
186	myb_DNA-binding	Myb-like DNA-binding domain	0.00011	29.1
187	pkinase	Eukaryotic protein kinase domain	3.4e-26	98.4
189	ER_lumen_recept	ER lumen protein retaining receptor	3.9e-144	492.2
190	ER_lumen_recept	ER lumen protein retaining receptor	2.1e-88	307.1
195	EMP24_GP25L	emp24/gp25L/p24 family	6.9e-06	28.1
199	zf-B_box	B-box zinc finger.	5.2e-07	36.7
211	HECT	HECT-domain (ubiquitin-transferase).	1.1e-115	397.8
213	Rhomboid	Rhomboid family	4.2e-42	153.3
214	LIM	LIM domain containing	8.8e-35	127.8

SEQ ID NO:	pFAM NAME	DESCRIPTION	p-value	pFAM SCORE
		proteins		
215	Ricin_B_lectin	Similarity to lectin domain of ricin	0.0015	19.2
216	chromo	'chromo' (CHRromatin Organization MOdifier	2.1e-09	37.1
218	Sialyltransf	Sialyltransferase family	7.3e-20	79.4
219	PG_binding_2	Putative peptidoglycan binding domain	5e-06	33.5
223	zf-C2H2	Zinc finger, C2H2 type	1.5e-104	360.7
226	RGS	Regulator of G protein signaling domain	5.1e-52	186.2
227	TBC	TBC domain	7.2e-35	129.3
228	CRAL_TRIO	CRAL/TRIO domain.	4.5e-47	158.6
232	Ribosomal_L44	Ribosomal protein L44	1e-48	175.3
235	ig	Immunoglobulin domain	3.5e-08	31.4
236	thyroglobulin_1	Thyroglobulin type-1 repeat	3.9e-24	93.6
238	TBC	TBC domain	1.2e-54	195.0
241	zf-C2H2	Zinc finger, C2H2 type	3.8e-08	40.5
242	AAA	ATPases associated with various cellular act	2.1e-43	157.6
249	integrin_A	Integrin alpha cytoplasmic region	0.091	18.0
256	PAP2	PAP2 superfamily	0.00084	22.8
257	zf-C2H2	Zinc finger, C2H2 type	1.2e-60	214.9
259	G-gamma	GGL domain	5.5e-30	108.3
266	efhand	EF hand	3.4e-07	37.4
267	efhand	EF hand	3.4e-07	37.4
274	zf-C2H2	Zinc finger, C2H2 type	0.00014	28.6
277	RecF	RecF protein	0.036	11.1
281	CH	Calponin homology (CH) domain	7.9e-22	86.0
285	cyclin	Cyclin	3.9e-07	28.5
289	zf-C2H2	Zinc finger, C2H2 type	1.9e-21	84.7
290	PI-PLC-X	Phosphatidylinositol-specific phospholipase	0.073	10.8
299	60s_ribosomal	60s Acidic ribosomal protein	4.1e-07	25.8
307	trypsin	Trypsin	6.9e-81	257.3
310	ig	Immunoglobulin domain	1.3e-10	39.3
311	ig	Immunoglobulin domain	6.1e-07	27.4
313	globin	Globin	3.8e-21	78.2
315	ig	Immunoglobulin domain	1.6e-05	22.8
318	zf-C2H2	Zinc finger, C2H2 type	9e-19	75.8
319	ig	Immunoglobulin domain	0.01	13.8
320	BTB	BTB/POZ domain	5e-17	70.0
322	aa_permeases	Amino acid permease	0.0058	-262.2
325	ig	Immunoglobulin domain	1.6e-10	38.9
327	ig	Immunoglobulin domain	1.9e-09	35.5
328	ig	Immunoglobulin domain	2.9e-09	34.9
329	ig	Immunoglobulin domain	7.4e-14	49.7
332	Ribosomal_L36e	Ribosomal protein L36e	6.3e-17	69.7
333	connexin	Connexin	7.6e-148	504.6
335	p450	Cytochrome P450	2.1e-100	347.0
337	Sm	Sm protein	0.00012	28.8
338	zf-C2H2	Zinc finger, C2H2 type	0.0025	24.5

SEQ ID NO:	pFAM NAME	DESCRIPTION	p-value	pFAM SCORE
343	aldo ket red	Aldo/keto reductase family	2.4e-53	190.7
345	ubiquitin	Ubiquitin family	3.1e-13	45.5
346	CH	Calponin homology (CH) domain	0.0017	23.8
351	ig	Immunoglobulin domain	4.8e-18	63.2
352	ig	Immunoglobulin domain	4.8e-18	63.2
358	G-alpha	G-protein alpha subunit	4.5e-148	505.3
359	ig	Immunoglobulin domain	8.9e-09	33.3
362	Ribosomal_L37ae	Ribosomal L37ae protein family	0.00083	-3.0

TABLE 5

SEQ ID NO:	POSITION OF SIGNAL IN AMINO ACID SEQUENCE	maxS (MAXIMUM SCORE)	meanS (MEAN SCORE)
2	1-29	0.942	0.664
12	1-15	0.909	0.589
14	1-17	0.974	0.943
20	1-22	0.932	0.802
25	1-16	0.988	0.881
28	1-13	0.896	0.771
37	1-21	0.992	0.929
42	1-46	0.978	0.754
52	1-34	0.954	0.756
63	1-31	0.960	0.773
71	1-45	0.981	0.652
80	1-22	0.982	0.882
81	1-42	0.993	0.715
83	1-30	0.966	0.767
95	1-18	0.997	0.971
102	1-13	0.981	0.764
107	1-45	0.890	0.631
110	1-27	0.992	0.969
138	1-33	0.961	0.864
144	1-45	0.987	0.658
145	1-20	0.992	0.967
175	1-20	0.957	0.874
176	1-21	0.989	0.945
179	1-42	0.980	0.577
184	1-20	0.972	0.771
189	1-28	0.941	0.755
190	1-28	0.941	0.755
191	1-12	0.907	0.779
195	1-21	0.958	0.779
200	1-15	0.970	0.875
211	1-20	0.895	0.595
215	1-31	0.987	0.895
218	1-30	0.971	0.889
225	1-17	0.884	0.588
235	1-23	0.965	0.817
237	1-29	0.933	0.725
249	1-28	0.972	0.870
251	1-17	0.966	0.905
260	1-26	0.921	0.587
270	1-20	0.938	0.631
283	1-18	0.901	0.763
288	1-20	0.940	0.693
293	1-26	0.937	0.784
295	1-22	0.972	0.745
296	1-15	0.930	0.748
297	1-35	0.906	0.600
300	1-29	0.981	0.864
307	1-19	0.976	0.916
308	1-27	0.973	0.931

SEQ ID NO:	POSITION OF SIGNAL IN AMINO ACID SEQUENCE	maxS (MAXIMUM SCORE)	meanS (MEAN SCORE)
309	1-29	0.950	0.629
310	1-19	0.969	0.913
311	1-21	0.956	0.823
315	1-17	0.976	0.938
317	1-19	0.943	0.837
319	1-18	0.991	0.978
324	1-26	0.968	0.806
325	1-20	0.972	0.828
326	1-27	0.893	0.567
327	1-21	0.994	0.959
328	1-20	0.945	0.891
329	1-21	0.984	0.858
330	1-27	0.891	0.593
333	1-40	0.955	0.703
347	1-22	0.968	0.806
351	1-23	0.982	0.945
352	1-23	0.982	0.945
355	1-32	0.955	0.617
356	1-23	0.936	0.677
359	1-20	0.937	0.859
360	1-29	0.956	0.765
361	1-23	0.968	0.819